

QRD-QLD searching based sphere detection for emerging MIMO downlink OFDM receivers

Predrag Radosavljevic*, Kyeong Jin Kim[†] and Joseph R. Cavallaro*

*Department of Electrical and Computer Engineering
Rice University, Houston, TX 77005, USA
{rpredrag, cavallar}@rice.edu

[†]Nokia Corporation, Irving, TX 75039 USA
{kyeong.j.kim}@nokia.com

¹ **Abstract**—In this paper, a detection algorithm with parallel partial candidate-search algorithm is presented. Two fully independent partial search processes are simultaneously employed for two groups of transmit antennas based on QR and QL decompositions of the channel matrix. Proposed QRD-QLD detection algorithm is compared with well-known QRD-M scheme adopted for several emerging wireless standards. Latency of the QRD-QLD candidate search is about twice as small for similar error-rate performance and for identical hardware resources. Total detection latency of QRD-QLD algorithm that also includes computation of soft information for outer decoder is also substantially smaller.

I. INTRODUCTION

Emerging wireless receivers are intended to support hundreds of Mbits/sec data-rates combined with excellent quality of service. In addition, high and flexible spectral efficiency is desired: multiple transmit antennas are accompanied with multiple receive antennas forming multiple-input multiple-output (MIMO) wireless transceivers [1]. The main challenge is to design high-throughput low-cost MIMO receivers that efficiently mitigate strong interference of different data streams transmitted from a plurality of antennas. Current practical receiver solutions employ detection based on minimum mean square error (MMSE) equalization combined with outer channel decoding such as low-density parity-check (LDPC) decoding [2]. The *a posteriori* probabilities (APPs) of transmitted coded bits may be transferred and improved between the inner detector and outer decoder. However, if the spectral efficiency of the system is high, the error-rate performance may still be far from the channel capacity.

In order to improve error-rate performance while achieving high spectral efficiency, approximation of exponentially complex maximum *a posteriori* (MAP) detection may be proposed at the receiver: soft sphere detection (SSD) algorithm proposed in [3]. This particular scheme may operate close to channel capacity if it is interfaced with outer soft decoding. However, the detection (or, equivalently, a search for valid transmission candidates) may still be too complex for efficient hardware implementation. Because of that, the design of sub-optimal SSD algorithms have been investigated, such as in [4], [5], [6].

¹This work was supported in part by Nokia Corporation and by NSF under grants CCF-0541363, CNS-0619767, CNS-0551692, CNS-0321266.

The latency of candidate search may also be large, especially if the number of transmit antennas and modulation size is high. It is, therefore, crucial to reduce computational complexity of the candidate-search algorithm without compromising the detection accuracy. In this work, the SSD with simplified candidate-search algorithm is proposed, while detection accuracy is preserved.

The paper is organized as follows. A MIMO downlink system with inner soft sphere detection and outer soft-input soft-output decoding at the mobile station is introduced in Section II. Simplification of the candidate-search process based on QR and QL decomposition of the MIMO channel matrix is described in Section III. Computational complexity, latency, and error-rate performance of the proposed QRD-QLD detection algorithm are compared in Section IV versus the well-known QRD-M detection algorithm [5] adopted for several emerging wireless standards. The paper is concluded in Section V.

II. MIMO-OFDM DOWNLINK SYSTEM AND SPHERE DETECTION

A coded MIMO downlink system based on Orthogonal Frequency Division Multiplexing (OFDM) transmission with soft inner detection and outer channel decoding at the mobile station (receiver side) is shown in Fig. 1. This system can be

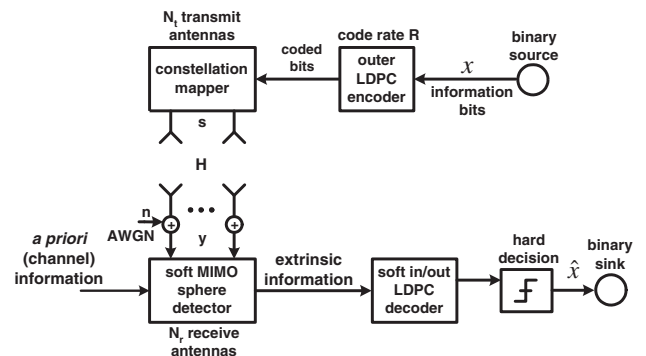


Fig. 1. MIMO transceiver, iterative detection and decoding.

defined by the linear equation:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (1)$$

where N_t and N_r are the number of transmit and receive antennas respectively, \mathbf{y} represents vector of N_r received symbols, \mathbf{H} is the $N_r \times N_t$ matrix of flat-fading channel coefficients that correspond to one frequency bin (or sub-carrier), \mathbf{s} is a vector of N_t transmitted modulated symbols, and \mathbf{n} is a vector of additive noise at the receive antennas.

The maximum-likelihood (ML) detection assumes testing of all possible transmit vectors Λ for the minimum square error cost given by $\hat{\mathbf{s}}_{ML} = \arg \min_{\mathbf{s} \in \Lambda} \|\mathbf{H}\mathbf{s} - \mathbf{y}\|^2$. The sphere detection represents simplification of the ML detection where tested candidates may be constrained to only those that are inside the hyper-sphere with the pre-determined radius r formed around the received symbol-vector \mathbf{y} [7]:

$$d(\mathbf{s}) = \|\mathbf{H}\mathbf{s} - \mathbf{y}\|^2 \leq r^2. \quad (2)$$

For the computationally simpler recursive checking, Eq. 2 may be transformed into an identical problem with triangular channel matrix after applying the QR decomposition: $\mathbf{H} = \mathbf{Q}\mathbf{R}$. Matrix \mathbf{R} is $N_t \times N_t$ upper triangular, and matrix \mathbf{Q} is a unitary $N_r \times N_t$ matrix. Equation 2 becomes $d(\mathbf{s}) = \|\mathbf{R}\mathbf{s} - \mathbf{Q}^H\mathbf{y}\|^2 \leq r^2$. Since matrix \mathbf{R} is upper triangular, the distance $d(\mathbf{s})$ can be calculated recursively from one transmit antenna to another:

$$T_m(\mathbf{s}) = T_{m+1}(\mathbf{s}) + \left| \hat{y}_m - \sum_{j=m}^{N_t} R_{mj}s_j \right|^2 \leq r^2, m = N_t, \dots, 1 \quad (3)$$

\hat{y}_m is a zero-forcing estimate of the signal transmitted from the m -th transmit antenna, R_{mj} is an entry of matrix \mathbf{R} that belongs to the m -th row and the j -th column, s_j is the symbol candidate for the j -th transmit antenna, and $T_i(\mathbf{s})$ is a partial Euclidian distance (PED) of candidate symbol \mathbf{s} at the search level m . $T_{M+1}(\mathbf{s}) = 0$ for all possible transmission vectors $\mathbf{s} \in \Lambda$. The SSD algorithm generates the list \mathcal{L} of candidates which provides reliability information for each transmitted coded bit. This information is passed as an extrinsic probability to the outer channel decoder. The extrinsic probabilities can be computed as in [3] using the max-log approximation.

We propose simplification of the candidate-search algorithm originally employed as a part of the bounded soft sphere detection (BSSD) from [8]. It is shown that the candidate-search algorithm can be divided into two independent parts for transmission candidates that correspond to two groups of antennas. Two partial search processes are simultaneously performed which decreases detection latency. The proposed QRD-QLD algorithm is compared with the QRD-M candidate-search from [5] which has been adopted as a possible solution for downlink OFDM receivers employed in several emerging wireless standards, such as the 3GPP-LTE, IMT-Advanced, WLAN, WiMAX, etc.

III. QRD-QLD CANDIDATE-SEARCH ALGORITHM

The QRD-QLD candidate-search algorithm is presented and compared with the QRD-M scheme in terms of computational complexity, latency and error-rate performance. Wireless OFDM system with four transmit/receive antennas is assumed,

but the proposed candidate-search algorithm can be applied for any symmetric odd or even number of transmit/receive antennas. It is assumed that the QRD-M detection starts from the most reliable antenna: antenna with the largest signal-to-noise ratio (SNR). At every search level, the best M candidates can be preserved. On the other hand, the main idea of the QRD-QLD detection is to simultaneously search for partial candidates from two groups of transmit antennas. The sub-carrier channel matrix is illustrated in Fig. 2 along with upper-triangular matrix \mathbf{R} and lower-triangular matrix \mathbf{L} obtained after QR and QL decomposition, respectively. For example, the 4-th transmit antenna may be the most reliable one, the 3-rd transmit antenna may be the second most reliable, while the 1-st antenna may be the least reliable. The partial search for symbol-candidates transmitted from the 4-th and 2-nd antennas may be performed simultaneously with the partial search for candidates transmitted from the 3-rd and 1-st antennas.

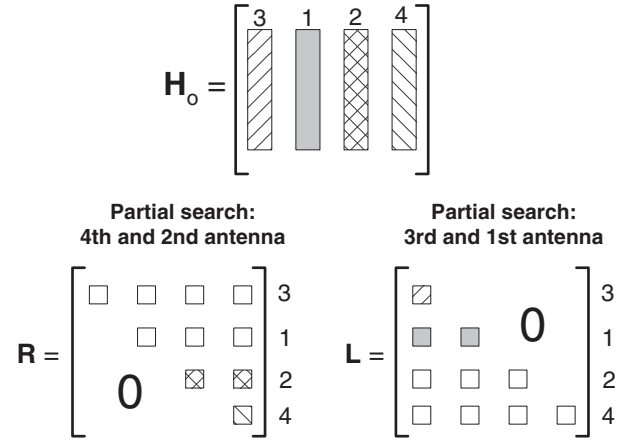


Fig. 2. Channel matrix \mathbf{H}_o after ordering of transmit antennas. Upper-triangular matrix \mathbf{R} and lower-triangular matrix \mathbf{L} are obtained after QRD and QLD of channel matrix \mathbf{H}_o . Two partial search processes are performed simultaneously. Colored elements of matrices \mathbf{R} and \mathbf{L} are used in the candidate-search process.

Instead of applying a QL-decomposition, an identical result can be obtained from QR-decomposition of the channel matrix with a reversed order of columns, as shown in Fig. 3. Colored elements of the upper-triangular matrix \mathbf{R}^{rev} in Fig. 3 are utilized in the partial search for candidates transmitted from the 3-rd and 1-st antennas. These elements are identical to elements of the lower-triangular matrix \mathbf{L} from Fig. 2.

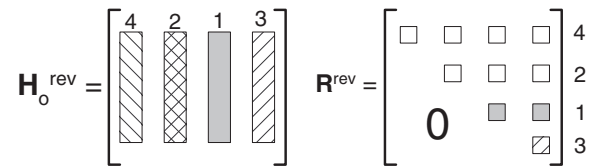


Fig. 3. Channel matrix with reversed order of channel-columns \mathbf{H}_o^{rev} , and corresponding upper-triangular matrix \mathbf{R}^{rev} obtained after QR-decomposition. Non-zero elements of matrix \mathbf{R}^{rev} are identical to non-zero elements of matrix \mathbf{L} from Fig. 2.

At the end of the partial search process for the 4-th and 2-nd transmit antennas, the partial Euclidian distance (PED) can be calculated for every transmission symbol-candidate s_2 for the 2-nd transmit antenna with the parent transmission candidate s_4 from the 4-th transmit antenna. The partial vector candidate $[s_4 \ s_2]$ is valid if it is inside the hyper-sphere with radius r :

$$T_2(\mathbf{s}_{42}) = T_4(s_4) + |\hat{y}_2 - R_{44}s_4 - R_{33}s_2 - R_{34}s_4|^2 \leq r^2, \quad (4)$$

where $\mathbf{s}_{42}=[s_4 \ s_2]$. Similarly, for the partial search process for the 3-rd and 1-st transmit antennas and according to Fig. 3, the partial vector candidate $[s_3 \ s_1]$ is valid if it is inside the hyper-sphere with radius r :

$$T_1(\mathbf{s}_{31}) = T_3(s_3) + |\hat{y}_1 - R_{44}^{rev}s_4 - R_{33}^{rev}s_2 - R_{34}^{rev}s_4|^2 \leq r^2, \quad (5)$$

where $\mathbf{s}_{31}=[s_3 \ s_1]$ is the partial vector-candidate for the 3-rd and 1-st transmit antennas.

In the case of the QRD-M approach from [5], a sorting of candidates is applied after every search level and up to M most reliable candidates can be preserved for further search. Therefore, up to M final candidates can be available at the end of the search process. On the other side, in the case of QRD-QLD detection, two simultaneous partial search processes are performed for two pairs of transmit antennas as shown in Fig. 4. The number of valid partial candidates is upper-bounded by $Cand$ candidates at the end of both partial search processes. Then, sorting of partial candidates can be applied and two groups of M most reliable partial candidates out of $Cand$ partial candidates are preserved. Both groups of partial candidates are then fully combined to generate the list of M^2 final candidates for all transmit antennas.

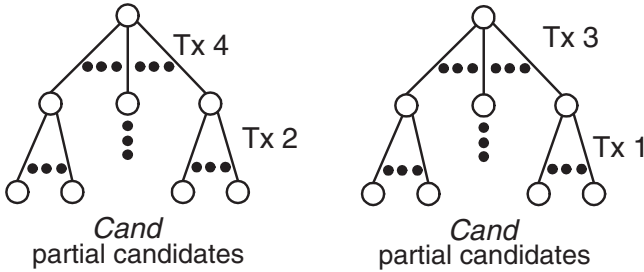


Fig. 4. Tree-search visualization for the QRD-QLD detection. Two parallel partial search algorithms generate two lists of partial candidates.

IV. COMPUTATIONAL COMPLEXITY, LATENCY, AND ERROR-RATE PERFORMANCE: QRD-QLD VS. QRD-M

Computational complexity, latency, and error-rate performance of the QRD-QLD algorithm are compared with QRD-M detection. Table I summarizes the arithmetic complexity of a single search operation, as well as complexity of the sorting algorithm applied to find the M smallest Euclidian distances (M most reliable candidates). The constellation symbol can be represented with M_C information bits.

The total number of arithmetic operations can be computed as a function of the parameter M and it is shown in Fig 5

TABLE I
NUMBER OF DIFFERENT ARITHMETIC OPERATIONS FOR QRD-M AND QRD-QLD CANDIDATE-SEARCH ALGORITHMS.

	sq.mul.	add	compare
search op.	$2\sqrt{2^{M_C}}$	$2^{M_C+1} + 2\sqrt{2^{M_C}}$	2^{M_C}
$sort_{QRD-M}$	-	-	$2^{M_C}(N_t - 1)M^2$
$sort_{QRD-QLD}$	-	-	$\frac{N_t}{2} \cdot M \cdot Cand$

for both QRD-M and QRD-QLD candidate-search algorithms. Also, Table II shows the number of arithmetic operations for values of parameters M and $Cand$ from Fig. 6. An addition operation is assumed to be a basic arithmetic operation. The number of square multiplications is scaled five times because it is typically five times more complex than addition for the same precision of input operands [9]. The number of comparisons is down-scaled by a factor of 2 since it is typically 50% faster and simpler than an addition operation for the same arithmetic precision. The QRD-QLD candidate-search algorithm is about half as computationally complex as the QRD-M candidate-search, except for smaller values of parameter M . As shown in Fig. 6, detection accuracy of the QRD-QLD algorithm with $M=18$ is very similar to accuracy of the QRD-M algorithm with $M=16$, while the computational complexity is twice as simple. The identical outer LDPC decoding is applied for both inner detection algorithms, codeword size is 1944 bits, code rate is 1/2, and there are 15 inner iterations of layered belief-propagation algorithm. There are up to four iterations between the outer decoder and inner detector.

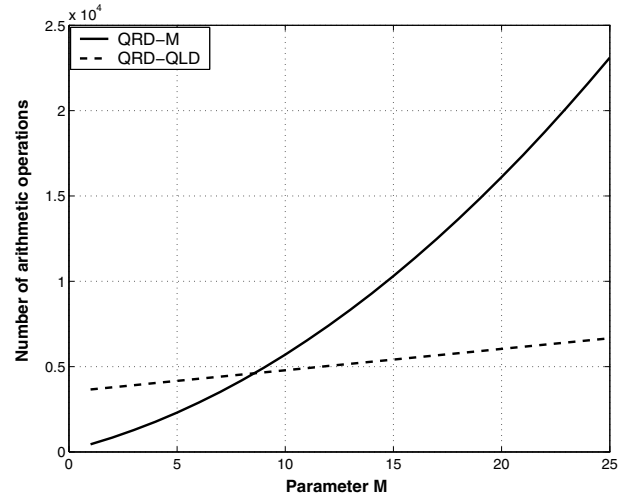


Fig. 5. Number of operations in the candidate-search process as a function of parameter M , 4x4 16-QAM Rayleigh fading channels: QRD-M vs. QRD-QLD, $Cand=125$.

A. Candidate-search latency of QRD-M and QRD-QLD algorithms

Latency of the QRD-M candidate-search algorithm is given by Eq. 6. The scaling factor of one half is applied because of

TABLE II
NUMBER OF ARITHMETIC OPERATIONS IN QRD-M AND QRD-QLD
CANDIDATE-SEARCH ALGORITHMS; 4x4 16-QAM SYSTEM.

	search process	candidate sorting	total number of arithm. ops.
QRD-M ($M=16$)	5096	6000	11153
QRD-QLD ($M=18$)	3536	2079	5672

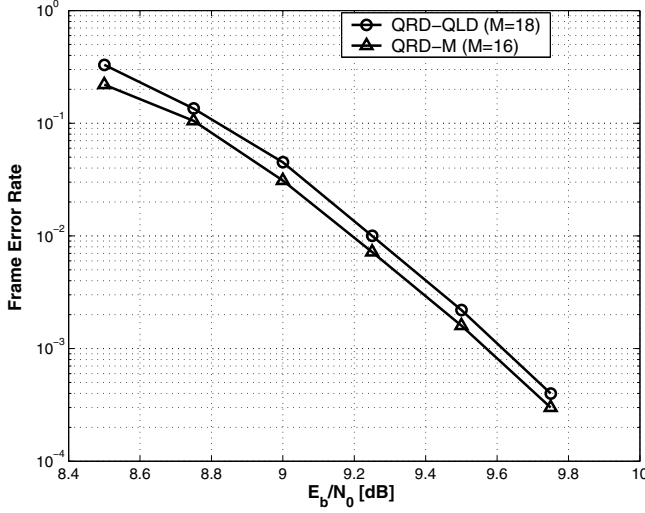


Fig. 6. Frame error rate performance in Rayleigh channels, 4x4 16-QAM: QRD-QLD with $M=18$ vs. QRD-M with $M=16$. Up to four outer iterations, LDPC decoding, code size of 1944 bits, code rate 1/2.

two search units that operates in parallel. As explained before, the QRD-QLD detection utilizes two parallel and independent search processes for two groups of transmit antennas, and therefore two candidate-search units are necessary to exploit parallelism of the QRD-QLD algorithm.

$$\begin{aligned}
 Lat_{..QRDM} &= \frac{1}{2} \cdot [1 + (N_t - 1) \cdot M] \cdot Lat_{..search_op}. \\
 &+ M \cdot [\alpha + M_C] \cdot Lat_{..compare_op}. \\
 &+ M(N_t - 1) \cdot (\beta + \lceil \log_2(M \cdot 2^{M_C}) \rceil) \cdot Lat_{..compare_op}, (6)
 \end{aligned}$$

where parameters α and β account for additional comparison delay in the candidate-sorting applied after the first and remaining N_t-1 search levels, respectively. These parameters are equal to zero if there are enough comparators to achieve the maximum level of processing parallelism for a given constellation size. Otherwise, they have non-zero values.

The QRD-QLD candidate-search latency including the sorting of candidates is approximately:

$$\begin{aligned}
 Lat_{..QRD_QLD} &= (1 + 2^{M_C}) \cdot Lat_{..search_op}. \\
 &+ M \cdot (\gamma + \lceil \log_2 Cand \rceil) \cdot Lat_{..compare_op}, (7)
 \end{aligned}$$

where $Cand$ is the maximum number of partial candidates preserved after the last search levels of two parallel search processes (the second search levels for $N_t=4$ transmit antennas), and parameter γ is similarly defined as parameters α

and β from Eq. 7. The value of parameter γ depends on the number of available comparators utilized for searching for the best M transmission candidates.

In the case of QRD-M candidate-search, the sorting of candidates is applied after every search level. After the first search level, the M best out of $P_C = 2^{M_C}$ candidates are preserved and used in the following search level. For the remaining N_t-1 search levels, the M best out of $M \cdot 2^{M_C}$ partial transmission candidates are saved. Two simultaneous candidate-sorting algorithms are applied at the end of both partial candidate-search processes, and the best M partial candidates can be determined out of $Cand$ candidates. Instead of sorting the entire list of candidates, only the M transmission candidates with the smallest Euclidian distances can be found. It is well known in the literature that the average latency of the conventional bubble-sort algorithm for a list of N unsorted elements can be approximately $N \cdot \log_2 N$ clock cycles. The sorting latency of the QRD-M algorithm can be reduced to $M \cdot (\xi + \log_2 N)$ clock cycles because it is sufficient to find only the M smallest elements. Parameter ξ accounts for additional latency if the number of available comparators cannot support full processing parallelism, which assumes enough comparators to find the smallest element out of N elements in up to $\lceil \log_2 N \rceil$ comparison stages.

Table III shows candidate-search latency of the proposed QRD-QLD detection versus the QRD-M detection. Two candidate-search units are available in both cases, as well as the same number of comparators. It can be observed that the candidate-search latency of the QRD-QLD algorithm is twice as small.

TABLE III
CANDIDATE-SEARCH LATENCY FOR QRD-M ($M=16$) AND QRD/QLD
($M=18$) AS A FUNCTION OF THE NUMBER OF COMPARATORS FOR
SORTING OF CANDIDATES, $Cand=125$ IN THE QRD-QLD DETECTION.

Algorithm and number of comparators	Search Latency [clks]
QRD-M ($M=16$, $\frac{1}{2} M \cdot 2^{M_C}=128$ comp.)	178
QRD/QLD ($M=18$, 128 comp.)	77
QRD-M ($M=16$, 64 comp.)	194
QRD/QLD ($M=18$, 64 comp.)	83
QRD-M ($M=16$, 32 comp.)	242
QRD/QLD ($M=18$, 32 comp.)	89

B. Latency for computing the soft information

In the QRD-QLD detection, the best M partial vector-candidates for two groups of transmit antennas are combined together forming a list of M^2 final vector-candidates. The latency for calculating the soft information for the outer decoder is obviously larger than in the case of the QRD-M algorithm where only M final vector-candidates are available. Latency for calculating the *a posteriori* probabilities (APPs) of coded bits for a single MIMO channel realization using one final vector-candidate is approximately one clock cycle [10]. Therefore, if P parallel APP function units are utilized and if there are C final vector-candidates, the latency of computing

final APPs is approximately C/P clock cycles. For the QRD-M detection with parameter $M=16$ and with $P=4$ parallel APP function units, the latency is approximately 4 clock cycles. In the case of QRD-QLD detection with $M=18$ (similar error-rate performance as QRD-M with $M=16$, see Fig. 6), the latency for calculating APPs is approximately 81 clock cycles.

C. Total detection latency and error-rate performance

Total detection latency of the QRD-QLD scheme is compared with total detection latency of the QRD-M scheme. Total detection latency is equal to a summation of candidate-search latency (see Eq. 7 and Eq. 6 for QRD-QLD and QRD-M, respectively) and latency for calculating soft information for the outer decoder. It can be assumed an implementation with two parallel search units, four APP units, and 64 comparators for the sorting of candidates. Current practical solutions for downlink OFDM receivers, as well as those proposed for emerging wireless standards, typically assume one iteration between inner detector and outer decoder. Total latency of the QRD-QLD detection is smaller than the total latency of the QRD-M detection if:

$$\left(83 + \frac{M_{QRD-QLD}^2}{4}\right) clks < \left(194 + \frac{M}{4}\right) clks, \quad (8)$$

where $M_{QRD-QLD}$ is a parameter in the QRD-QLD candidate-search algorithm, and the latency of candidate-search for 64 comparators is available in Table III. If the QRD-M detection with $M=16$ is assumed as a reference, then the inequality (8) is satisfied for $M_{QRD-QLD} < 21.4$. Total detection latency assumes detection of one channel realization including computation of reliability messages for transmitted coded bits.

Figure 7 shows the frame error-rate performance of the QRD-M detection with $M=16$, and the QRD-QLD detection with two groups of $M=18$ most reliable partial candidates combined together forming $M^2=324$ final transmission candidates. One iteration between inner detector and outer decoder is considered. It can be observed that the frame error-rate performance of these two schemes are very similar. At the same time, the maximum total detection latency of the QRD-QLD scheme is about 18% lower compared to the latency of QRD-M detection with $M=16$ (162 clock cycles for the QRD-QLD detection versus 198 clock cycles for the QRD-M detection).

V. CONCLUSION

We present a partial candidate-search algorithm for soft sphere detection based on QR and QL decompositions of the channel matrix. Two independent and parallel search algorithms are applied for two separate groups of transmit antennas. The final list of candidates is obtained after combining two groups of independent partial candidates. The proposed novel QRD-QLD detection is compared in terms of detection latency and error-rate performance with the well-known QRD-M algorithm adopted for several emerging wireless standards. The search latency is twice as small for similar error-rate

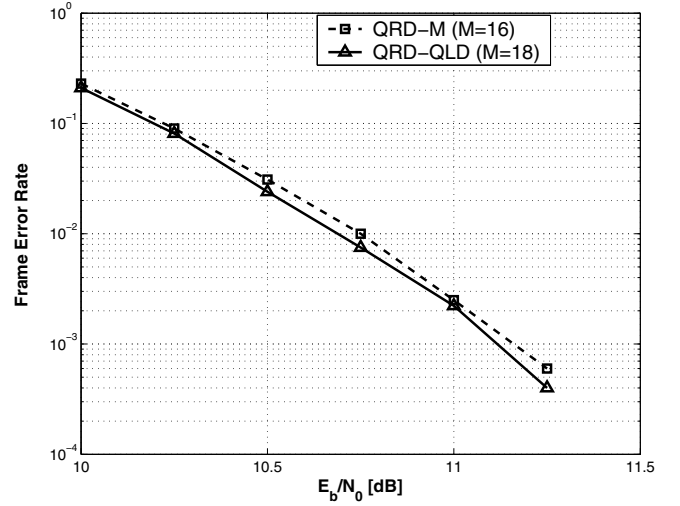


Fig. 7. Frame error rate performance in Rayleigh channels, 4x4 16-QAM: QRD-M ($M=16$) vs. QRD-QLD ($M=18$). One outer iteration, LDPC decoding, code size of 1944 bits, code rate 1/2, 15 inner iterations of LBP algorithm.

performance and for identical hardware complexity (same number of search units and comparators for candidate-sorting). Total detection latency of the QRD-QLD algorithm is about 18 % less for the same error-rate performance and identical hardware resources.

REFERENCES

- [1] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, pp. 41–59, 1996.
- [2] B. Lu, G. Yue, and X. Wang, "Performance analysis and design optimization of LDPC-coded MIMO OFDM systems," *IEEE Transactions on Signal Processing* [see also *IEEE Transactions on Acoustics, Speech, and Signal Processing*], vol. 52, pp. 348–361, Feb. 2004.
- [3] B.M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, pp. 389–399, March 2003.
- [4] D. Garrett, L. Davis, S. ten Brink, B. Hochwald, and G. Knagge, "Silicon complexity for maximum likelihood MIMO detection using spherical decoding," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1544–1552, Sept. 2004.
- [5] K. J. Kim, J. Yue, R.A. Iltis, and J.D. Gibson, "A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 710–721, March 2005.
- [6] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 491–503, March 2006.
- [7] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.
- [8] P. Radosavljevic and J.R. Cavallaro, "Soft sphere detection with bounded search for high-throughput MIMO receivers," in *Conference Record of the Fourtieth Asilomar Conference on Signals, Systems and Computers*, pp. 1175–1179, Oct.-Nov. 2006.
- [9] M.M. Mansour and N.R. Shanbhag, "High-throughput LDPC decoders," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, pp. 976–996, Dec. 2003.
- [10] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bolcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 1566–1577, July 2005.